IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:	)	
	)	
Guojun Zhou	)	Examiner: Abul K. Azad
	)	
Application No.: 09/685,419	)	Art Unit: 2654
	)	
	)	
Filed: 10/10/2000	)	**RECEIVED**
	)	
For: Language Independent Voice-based	)	AUG 1 2 2004
Search Engine	)	
	)	Technology Center 2600

DECLARATION UNDER 37 C.F.R. 1.132

I, Yonghong Yan, do hereby state that:

1. I am a senior researcher of the Institute of Acoustics of the Chinese Academy of Sciences. My research area is speech and natural language processing including speech recognition, speech synthesis, language identification, natural language based human-machine interaction, etc. I have been working in this field for more than 15 years. I obtained my Ph.D. degree in speech and natural language processing from the Oregon Graduate Institute (OGI) School of Science & Engineering of Oregon Health & Science University in 1995. Since then, I have worked as an assistant professor and an associate professor of the OGI School of Science & Engineering of Oregon Health & Science University until June, 2004. During this period of time, three people successfully completed their PhD study under my supervision, all in the speech and natural language processing field. From 1998 to 2001, I directed a research lab for Intel Corporation to work on improving the performance of speech and natural language processing applications using Intel Architecture-based computing systems. Additionally, I have published over 20 research papers in my research field in the U.S. and internationally.

2. In my opinion, prosodic pattern processing is not inherent in a text-to-speech (speech synthesis) system. This is why some text-to-speech systems produce very mechanical (machine-like) speech. Systems with prosodic patterns sound much more natural than those without prosodic patterns. How to generate and add a prosodic pattern to synthesized speech to make it sound natural has been a very challenging research topic in the speech and natural language processing field. In the past 3-4 years, some progress has been made and now it is relatively easier than before to find a commercial text-to-speech system that can produce speech that sounds closer to naturally spoken human speech. This capability was not well known in October, 2000.

3. In the 2000 timeframe of the date of the above-identified application, researchers in the speech and natural language processing field were trying to build commercial natural spoken language based dialogue human-machine interaction applications. Those efforts were focused on relatively simple applications (e.g., spoken language driven television program guide) where users' speaking styles are relatively predictable, the background database of recognizable speech terms (e.g., an electronic television program database) is limited, and responses to a user's requests were relatively simple. Little efforts were made in more complicated applications (e.g., spoken language based human interactions with the world wide web) where users' speaking styles are more natural (e.g., a user may start speaking one thing but suddenly change to something else; or a user may start speaking one thing, but insert something unrelated within a sentence, etc.), background databases are very complex, and responses to a user's request may be so complicated to involve some kind of simplification process. Even if there were such efforts, they were very unlikely to succeed. One reason for this is that most speech recognition systems were not sophisticated

enough to recognize any naturally spoken language (e.g., those with filler words, repeats, insertions, etc.). Another reason is that naturally spoken language processing was very computationally intensive and it was very difficult to achieve real-time performance with then available computing systems. Additionally, it was unpredictable then whether these technical difficulties could be solved in the near future.

4. Regarding automatic summarization technologies, at the time of invention in mid-2000 they were mainly used to summarize a single document. A couple of years ago (2002), researchers started to make these technologies work for multiple documents and have achieved some extent of success. Such success is possible mainly because technical progress made in natural language processing has been used for automatic summarization and because the speed of processors makes the time required to summarize multiple documents relatively acceptable. However, at the time of invention a working automatic summarization capability (especially for search results from the Internet) was not well-known.

5. In the Internet search domain, it was hard to imagine that someone could build a language independent voice-base search engine before the year of 2001. Each component involved (e.g., speech recognition, natural language processing, automatic summarization, etc.) is very computationally intensive. Achieving real-time performance with even a single component was a big challenge in the late 1990s and 2000, let alone achieving real-time performance with a combination of several such components. Additionally, technologies involved in each component were not sophisticated enough to make it predictable to build a workable language independent voice-based search engine then. With the advance in both technologies and processor speed in the year of 2000 and later years, an application of a language independent voice-based

search engine may be built, although continuing improvements in processor speed and technologies may make it work much better in the future.

Respectfully submitted,

Dated: 7/8/04

Yonghong Yan